Providing Access to Environmental Data in a SOA in NOKIS

Carsten Heidmann¹, Frank Sellerhoff², Rainer Lehfeldt, Aleksei Valikov³ and Wassilios Kazakos

Abstract

NOKIS and NOKIS++ are two German R&D projects, which aim at the integration of distributed data from different sources provided in various formats. The authors describe a Service Oriented Architecture, which has been implemented for users from coastal zone management. The available data types are characterized and the categories of used services are explained, together with the relevant standards in their respective sectors of application. Examples of the implemented services are given and the instrument to combine the available services is introduced. The article finishes with a short comparison between a SOA and central approaches.

1. Project Background

The authors have gathered knowledge about the access to and the visualization of environmental data in two projects. NOKIS, the "North Sea and Baltic Sea Coastal Information System" (2001-2004) and NOKIS++, "Information Infrastructures for the North Sea and Baltic Sea Coast as Contribution to Integrated Coastal Zone Management" (2004-2008) are research projects, funded by the German Federal Ministry of Education and Research (Lehfeldt 2002, Heidmann 2004). Their goal is the documentation of research results and monitoring data from coastal authorities and research institutions with the help of a common metadata standard and the access to this data for further computing and visualization purposes.

The main problems to be tackled were the missing or incomplete documentation and above all the problems related to different data formats and network security measures. NOKIS provides a unified metadata infrastructure for the participating partners, combined with tools to generate, maintain and distribute metadata.

2. Data Types in NOKIS

Project partners from a variety of research institutions and state agencies cooperated in NOKIS. As a result, the central metadata repository includes documentation about a wide range of data types commonly used. Examples include GIS data from the national park offices, containing monitoring results from the German Wadden Sea as well as extensive time series from tidal gauges which are maintained by coastal protection authorities on state level.

¹ Federal Waterways Engineering and Research Institute, Wedeler Landstrasse 157, 22559 Hamburg, Germany, {carsten.heidmann|rainer.lehfeldt}@baw.de, http://www.baw.de
² smile consult GmbH, Vahrenwalder Strasse 7, 30165 Hannover, Germany, sellerho@smileconsult.de, http://www.smileconsult.de
3. Using Metadata to Control Information Flow

The central paradigm for creating the NOKIS service infrastructure is to avoid unnecessary hard-wired connections between data and services. All information should be retrieved from metadata repositories, maintained either centrally or locally. This information should be used to build working environments on-the-fly, thereby integrating all available services at the time the user invokes the initial search.

4. Service Types in NOKIS

4.1 Relevant Standards

Another paradigm is the use of widely acknowledged international standards, which provide a solid basis for the implementation of the desired services. Since all of the data has a geospatial context, the standards of the Open Geospatial Consortium (OGC) are of special interest. For the usage in NOKIS, the Catalogue Service Web (CS-W, Nebert et al. 2007), the Web Mapping Service (WMS, Beaujardière 2002) and the Web Feature Service (WFS, Vretanos 2005) have been of greatest interest.

One of the first groups discussing standards for web services was the World Wide Web Consortium (W3C). Today, a lot of web services are using the SOAP protocol (Simple Object Access Protocol, Gudgin et al. 2007) and describe themselves with WSDL (Web Services Description Language, Christensen et al. 2001) documents. The OGC standards, however, adjust only slowly to these standards, which makes it difficult to include them in infrastructures which are using W3C compliant web services.

Another group of standards which are relevant for NOKIS is the suite of standards for geoinformation of the International Organization for Standardization (ISO). The most important ones for NOKIS are ISO 19115: Geographic information – Metadata (ISO 2003²), ISO 19136: Geographic information -- Geography Markup Language (ISO 2007), and ISO 19139: Geographic information -- Metadata -- XML schema implementation (ISO 2007²).

4.2 Catalogue Services

Catalogue Services are the central component in the NOKIS infrastructure. The catalogues hold information about the data proper, about data access services, and about the services available for visualization and analysis.

The catalogue services provide standard-compliant metadata to serve higher level metadata information systems in Germany like Geoportal.Bund of the GDI-DE, PortalU for environmental information and in the future INSPIRE. Every single NOKIS instance has a built-in CSW interface, which provides geo-metadata compliant to the German ISO19115-profile known as “DE-Profil” (Voges 2005).

The additional information about services for internal use in the NOKIS infrastructure is also exchanged with an enriched metadata profile, which is a superset of the DE-Profil. In this way, it is possible to satisfy external as well as internal requests for metadata by using the standard CSW interface.

4.3 Gazetteer Service

Environmental data normally has a spatial footprint. A lot of data retrieval systems provide the possibility to focus the search on a user defined bounding box. This is not always satisfying, because the user often is familiar with precise geographic names without knowing exactly where to put a bounding box.
Fig. 1: Example for Gazetteer Data - Multiple Geometries for the Island Trischen

A digital gazetteer provides the user with options to run a spatial query without defining a spatial extent on his own. The entered search terms are checked against the gazetteer vocabulary and all existing geometries are used as constraints in the following search.

A gazetteer for the German coast has been developed in NOKIS, which provides place names and geometries for the coastal zone (Kohlus 2007). The data for this gazetteer has been compiled from a variety of sources, including maps of geographical names (StAGN 2004), official geodata (ATKIS®, AdV 2008) and personal communications. The NOKIS gazetteer also covers the temporal variability of geometries and names, which is normally not included in conventional gazetteers but which is of special interest for the tidal flats along the German coast. The speed of morphological changes in this region is very high and it is not unusual to observe local displacements of morphological features in a magnitude of several meters within a few days. Since the German coast is a region where different languages have been used during the course of time, many places have more than one name, either because it changed with the time or because there are names in different languages. The NOKIS gazetteer models this variability in its data model, which is a substantial extension of the ADL gazetteer model (ADL Project 2004).

The gazetteer itself is accessible as a WFS service, which allows the inclusion in standard OGC compliant clients. There is ongoing work to provide a WFS, which is compliant to the ISO 19112 - Spatial referencing by geographic identifiers (ISO 2003).

4.4 Data Access Services

Metadata systems provide a way to find information about data. For the integration in a Service Oriented Architecture (SOA), it is necessary to provide ways to access the information itself. The WFS specification allows to access vector geodata. Coastal data, however, consist of many other data types, which cannot simply be accessed via WFS, or where the focus is concentrated on information other than the geome-
try. Examples for those data types are time series of wave height and direction or current fields as result of numerical modelling. It was the aim to keep the coupling of services in NOKIS as loose as possible, so that any data set can be used with the greatest number of services possible. Therefore, an object model has been created, which covers all data types in NOKIS. This model allows determining the return objects of the different data access services at runtime (late binding), avoiding hardcoded data types in the created W3C web services.

### 4.5 Visualization Services

Visualization services in NOKIS focus on the visualization of time series and field data, which is not or only insufficiently covered by the OGC services. Especially the visualization of vector data are insufficiently covered. In coastal engineering, visualizations of vector field data (e.g. wind fields or current fields) and vector point data (e.g. wind direction and speed at a certain point, Figure 2) are frequently used. The visualization services in NOKIS are using the same data model as the data access services. This allows using the output objects of the data access services as input for the visualization services.

![Fig. 2: Distribution of Wind Direction and Speed](image)

### 4.6 Documentation of Data and Services

Based on the standards for ISO 19115 (ISO 2003) for metadata about geodata and ISO 19119 (ISO 2005) for metadata about spatial services several metadata profiles where defined to document the different data types used in NOKIS. The metadata sets are ISO compliant extensions of the standards.

A major challenge is to capture as much information about data and services as possible with these profiles. Since the personnel for data documentation is not necessarily familiar with the ISO standards, an easy to use editor had to be developed, providing an intuitive user interface, which hides most of the complexities of the standard’s data model.
The editor was developed in several iterations always incorporating new user feedback and suggestions and is now available in version 2.0. Although it is a pure web application, the use of web 2.0 technologies, makes the editor behave like a desktop-application, including pull down menus, tabs etc.

![NOKIS Editor](image)

Fig. 3: NOKIS Editor with validation information

As the quality of the metadata is of high importance for information portals and workflows depending on metadata, several quality assurance measures are implemented during both the editing and publication process. The former was realized by real-time validation of the metadata entries. The user immediately sees where new entries have to be made. Additionally the user is directly informed in case of wrong entries. In order to support quality control during the publication process, several sharing and access control mechanisms where introduced. The 4-eyes-principle ensures that the entries of the editor can be controlled by a super user before publication.

4.7 Service Orchestration in KFKI-GIS

Putting this infrastructure to work requires a platform, which is capable of invoking the different services and which knows how to combine the different services according to the user's needs. Since there are only few approaches of generic tools to invoke web services and no tools which allow a service chaining on-the-fly, it was necessary to develop a tool from scratch.
First attempts had been made with a Java Server Faces (JSF, Burns 2006) frontend. They showed that it is necessary to have a versatile Graphical User Interface, which is easier to implement than a Java Desktop Program using Swing as presentation layer. To allow a web based approach to the NOKIS service infrastructure, it was decided to wrap the application with Java Web Start. The most recent version is build around a map based interface, providing access to OGC web services as well as W3C web services which form the base of the NOKIS service architecture (Figure 4). This application serves as central tool for data integration of the German Coastal engineering Research Council (KFKI-GIS),

![Figure 4: KFKI-GIS user interface (map view)](image)

It includes a number of predefined OGC services (WMS and WFS), which allow a quick orientation within the area of interest. A special case is the use of the NOKIS Gazetteer Service, which is accessible directly and which shows the existing place names in the area of interest. Besides this, there are no hard-coded references between metadata and services; the invocation of data access services depends on the data type, documented in the metadata. They return an object, whose type is determined at runtime and the subsequent invocation of analysis and visualization services depends completely on the data types returned by data access services.
5. **Comparison with Central Approaches**

The use of a SOA has many advantages over a monolithic stand-alone application. First of all, the inventory of available services is always up to date. A new service or new data sets are immediately available to all NOKIS partners at the time of publication in one of the connected metadata catalogues. With the chosen architecture, it is possible to use common analysis and visualization services from a central repository and to add certain methods which are of importance to local users.

The SOA pattern, however, has a major disadvantage in comparison to stand-alone applications. The whole system may become unusable as soon as one of the central services is not available (like the central metadata repository). As long as there are no High Availability (HA) additions, this situation is likely to happen every now and then. Even with HA solutions, local network problems may also break the chain of necessary services.

Another aspect is quality management. A central data warehouse solution is more likely to provide a unified quality assurance procedure than a solution of distributed data sources. Here it is likely, that the quality of the different data sources can vary considerably.

6. **Conclusions and Outlook**

We have learned from this project, that a complex and dynamic system like NOKIS will only be accepted for the daily work, when there is an immediate benefit for the user who has provided data and metadata. By using a loose coupling between data and services, the expenses for maintenance and administration could be reduced significantly.

Data provider and user can access a working environment, which meets the needs for consistent documentation and which is continuously growing in functionality and scope. The documentation can be reused in a variety of reporting obligations. This effect can only be achieved by firstly making use of all relevant service standards, not only the common OGC standards like CSW, WFS, and WMS, and secondly by providing sufficient strictly standardized metadata for all system components.

Following these guidelines can help avoided creating an isolated application and to build a system, which is interoperable with other systems, thereby ensuring that the initial investment will not be lost due to changes in technology or exchange formats.

**References**


